Recruitment and remodeling of an ancient gene regulatory network during land plant evolution

Pires, Nuno D ; Yi, Keke ; Breuninger, Holger ; Catarino, Bruno ; Menand, Benoît ; Dolan, Liam

Abstract: The evolution of multicellular organisms was made possible by the evolution of underlying gene regulatory networks. In animals, the core of gene regulatory networks consists of kernels, stable subnetworks of transcription factors that are highly conserved in distantly related species. However, in plants it is not clear when and how kernels evolved. We show here that RSL (ROOT HAIR DEFECTIVE SIX-LIKE) transcription factors form an ancient land plant kernel controlling caulonema differentiation in the moss Physcomitrella patens and root hair development in the flowering plant Arabidopsis thaliana. Phylogenetic analyses suggest that RSL proteins evolved in aquatic charophyte algae or in early land plants, and have been conserved throughout land plant radiation. Genetic and transcriptional analyses in loss of function A. thaliana and P. patens mutants suggest that the transcriptional interactions in the RSL kernel were remodeled and became more hierarchical during the evolution of vascular plants. We predict that other gene regulatory networks that control development in derived groups of plants may have originated in the earliest land plants or in their ancestors, the Charophycean algae.

DOI: https://doi.org/10.1073/pnas.1305457110

CLASSIFICATION: Biological Sciences; Plant Biology

Recruitment and remodelling of an ancient gene regulatory network during land plant evolution

SHORT TITLE: Evolution of an ancient plant bHLH network

Nuno D. Pires[a,b], Keke Yi[b,c], Holger Breuninger[a], Bruno Catarino[a], Benoît Menand[b,d] and Liam Dolan[a,b,1]

[a]Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

[b]Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK

[c]Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

[d]Laboratoire de Génétique et Biophysique des Plantes iBEB, CNRS, CEA, Aix-Marseille Université, Marseille, F-13009, France

[1]To whom correspondence should be addressed:

Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK

liam.dolan@plants.ox.ac.uk

Tel +44 (0)1865 275147

Fax +44 (0)1865 275047

Abstract

The evolution of multicellular organisms was made possible by the evolution of underlying gene regulatory networks (GRNs). In animals, the core of GRNs consists of kernels, stable subnetworks of transcription factors that are highly conserved in distantly related species. However, in plants it is not clear when and how kernels evolved. We show here that RSL transcription factors form an ancient land plant kernel controlling caulonema differentiation in the moss *Physcomitrella patens* and root hair development in the flowering plant *Arabidopsis thaliana*. Phylogenetic analyses suggest that RSL proteins evolved in aquatic charophyte algae or in early land plants and have been conserved throughout land plant radiation. Genetic and transcriptional analyses in loss of function *A. thaliana* and *P. patens* mutants suggest that the transcriptional interactions in the RSL kernel were remodelled and became more hierarchical during the evolution of vascular plants. We predict that other gene regulatory networks that control development in derived groups of plants may have originated in the earliest land plants or in their ancestors, the Charophycean algae.

Introduction

The development of multicellular organisms is controlled by gene regulatory networks (GRNs) and the reorganisation of GRN architecture is thought to be a major factor underlying morphological evolution (1–5). GRN are hierarchic and modular structures where four major component classes can be identified (3): at the periphery of GRNs are differentiation gene batteries encoding proteins that execute cell type-specific functions (such as building a pigmented cell); upstream of these batteries are switches that allow or prevent subcircuits to function in specific developmental contexts, and 'plug-ins', small subcircuits that are flexibly and repeatedly used during development (such as signal transduction pathways); at the core of GRNs are kernels, small conserved subcircuits that execute specific developmental functions (such as defining spatial patterns in an embryo). Kernels are comprised of transcription factors that are highly conserved in distantly related species and are unusually stable components of GRNs.

Ancient kernels that regulate body plan and organ development are highly conserved among diverse groups of metazoans (animals) (6–10). By contrast, the core components of plant GRNs are difficult to identify because of the dynamic nature of plant genome evolution and the plastic character of plant development. Floral homeotic genes form a relatively recent kernel controlling flower development (11). Homologues of KNOX and LEAFY transcription factors control shoot development in vascular plants and sporophyte development in mosses (12–15). KNOX/BEL genes also control the development of the diploid phase in unicellular chlorophytes (16) and the haploid-to-diploid transition in mosses (15), suggesting that KNOX and LEAFY genes may be core members of ancient GRNs that control diploid development in plants. Auxin signalling (17, 18), ethylene perception (19), abscisic acid signalling (20) and several small RNAs are conserved between mosses and flowering plants (21, 22), suggesting that many switches and plugins of land plant GRNs have been conserved since before the evolution of vascular

plants over 440 million years ago. However, the architecture and evolutionary history of these hypothetical ancient GRN kernels is mostly unknown.

In the angiosperm *Arabidopsis thaliana*, root hair development is controlled by the bHLH transcription factors AtRHD6 (*Arabidopsis thaliana* ROOT HAIR DEFECTIVE 6) and AtRSL1 (*A. thaliana* RHD SIX-LIKE 1); their homologues in the moss *Physcomitrella patens*, PpRSL1 and PpRSL2, control the development of filamentous rooting structures: caulonema and rhizoids (23–25). This suggests that *RSL* genes belong to an ancient land plant GRN that controls the differentiation of cells with a rooting function. In *A. thaliana*, AtRHD6 was also found to form a transcriptional mechanism with two other bHLH transcription factors, AtRSL2 and AtRSL4 (26).

Here we test the hypothesis that the RSL mechanism is an ancient land plant kernel. We show that *RSL* genes form a transcriptional network that controls root hair development in *A. thaliana* and protonema development in *P. patens*. *RSL* genes form two ancient lineages that evolved in charophyte algae or in the first land plants and have been conserved during land plant evolution. Functional and expression analysis of the *RSL* genes in *A. thaliana* and in *P. patens* indicate that the two lineages form a transcriptional regulatory network in both species. Together, our results suggest that the *RSL* genes form a kernel that evolved over 450 million years ago and was recruited to control the development of root hairs during the evolution of vascular plants.

Results

The RSL network controls root hair development in *A. thaliana*

The differentiation of root hairs in *A. thaliana* is controlled by a regulatory mechanism that comprises the bHLH transcription factors AtRHD6, AtRSL1, AtRSL2 and AtRSL4: no root hairs differentiate in *Atrhd6 Atrsl1* or in *Atrsl2 Atrsl4* double mutants (23, 26) and the transcription of *AtRSL2* and *AtRSL4* is positively regulated by AtRHD6 and AtRSL1 (26). These four genes

belong to a phylogenetic group that also includes *AtRSL3* and *AtRSL5* (Fig. 1B). To determine if AtRSL3 and AtRSL5 also control root hair development, we characterised the phenotypes of *Atrsl3*, *Atrsl5* and *Atrsl2 Atrsl3* mutants. Root hairs of *Atrsl3* and *Atrsl5* single mutants were indistinguishable from wild type, but the root hairs of *Atrsl2 Atrsl3* were shorter than in *Atrsl2* single mutants (Fig. 1A). Furthermore, the constitutive expression of *AtRSL3* or *AtRSL5* in the hairless *Atrsl2 Atrsl4* double mutant background could partially restore root hair development (Fig. S1). Together, these data indicate that each of the six *A. thaliana RSL* genes positively regulate root hair development. *AtRHD6* and *AtRSL1* genes were expressed early in the development of trichoblasts (the epidermal cells that give rise to root hair cells), but the expression disappeared before root hairs initiate (23) (Fig. 1C). In contrast, *AtRSL2* and *AtRSL4* were expressed later, specifically during root hair growth (26) (Fig. 1C). Since *AtRSL3* and *AtRSL5* also control root hair growth, we hypothesised that they would be expressed while root hairs elongate. Accordingly, AtRSL3:GFP and AtRSL5:GFP protein fusions expressed under the control of their respective native promoters accumulated in the nuclei of trichoblasts during root hair growth (Fig. 1C). Together, this indicates that *AtRHD6* and *AtRSL1* act earlier in root hair development than *AtRSL2*, *AtRSL3*, *AtRSL4* and *AtRSL5*. Together, these results suggest that all six *RSL* genes are components of a transcriptional network that controls root hair development in *A. thaliana*.

RSL class I and class II genes were present in early land plants

We hypothesised that RSL genes form a GRN kernel that is present in other land plants. To trace the evolutionary history of the RSL regulatory network and define the diversity of *RSL* genes in plants, we identified and retrieved RSL sequences from 12 different plant genomes. RSL proteins are characterised by a conserved C-terminal region, which includes a bHLH domain that extends into a conserved stretch of 14 amino acids, the RSL domain (Fig. 2A).

Maximum likelihood phylogenetic analyses using RSL sequences from different species show that RSL proteins form two distinct and ancient phylogenetic clades, which we named RSL class I and RSL class II (Fig. 2B and Fig. S2; subfamilies VIIIc(1) and VIIIc(2) in (27). We found both RSL classes in all species of land plants for which a complete genomic sequence is available, including mosses, lycophytes, eudicots and monocots (Fig. 2C), but we did not find RSL sequences in chlorophyte algae. This indicates that, like most other plant bHLH subfamilies (27), RSL class I and class II proteins evolved sometime after the divergence of the chlorophyte and streptophyte lineages 700-1000 million years ago (28, 29), but before the evolution of vascular plants over 443 million years ago (30). This means that RSL proteins evolved in multicellular streptophytes either before or shortly after their colonisation of terrestrial environments.

The similarity of the amino acid sequences in the C-terminal region of RSL proteins in different land plants suggests that their molecular function may be conserved. To test this hypothesis, we transformed the hairless *A. thaliana Atrsl2 Atrsl4* (RSL class II) double mutant with *P. patens RSL* class II genes under the control of the constitutive cauliflower mosaic virus (CaMV) 35S promoter. Each of the moss genes could partially rescue the development of root hairs in *Atrsl2 Atrsl4* double mutant plants (Fig. 2D). The partial rescue ranged from the formation of small bulges in *Atrsl2 Atrsl4* plants expressing *PpRSL3* or *PpRSL4* to the development of short tip growing root hairs in plants expressing *PpRSL5* or *PpRSL6*. Similarly, the expression of the *A. thaliana RSL* class II genes *AtRSL3* and *AtRSL5* in *Atrsl2 Atrsl4* plants could partially rescue the development of root hairs (Fig. S1). Previously, Menand et al. (23) showed that a *P. patens* RSL class I protein could rescue the development of root hairs in the hairless *A. thaliana* RSL class I *Atrhd6 Atrsl1* mutant. Together, these data indicate that the molecular function of RSL class I and RSL class II proteins is conserved between mosses and angiosperms.

The RSL network controls protonema development in *P. patens*

The conservation of RSL protein function in land plants supports the hypothesis that *RSL* genes are components of an ancient GRN that was present in early land plants. To further test this hypothesis, we determined if an *RSL* gene regulatory network exists in the moss *P. patens*; if the RSL network is present in both *P. patens* and *A. thaliana* then it would also have been present in their common ancestor – an early land plant. *P. patens RSL* class I genes control the chloronema-to-caulonema transition in the moss protonema (23). The protonema is a filamentous structure composed of chloronema and caulonema cells: chloronema are small, slow growing cells with numerous chloroplasts that fulfil a predominantly assimilatory function; caulonema are larger, fast growing cells that play an important role in substrate colonisation (31). Young protonemata predominantly comprises cells with chloronema characteristics, but in older protonema several chloronema cells undergo a transition to caulonema. However, the differentiation of caulonema cells was totally abolished in the RSL class I *Pprsl1 Pprsl2* double mutant (23). If *RSL* class II genes function in the same pathway as *RSL* class I genes, we would expect that the chloronema-to-caulonema transition would also be defective in *RSL* class II mutants. To test this hypothesis, we generated *P. patens* plants that lack the function of single or paralogous pairs of *RSL* class II genes. *Pprsl3 Pprsl4* double mutants developed small and very dense protonemata composed predominantly of chloronema cells (Fig. 3), indicating that the chloronema-to-caulonema transition is defective. Unlike the *Pprsl1 Pprsl2* RSL class I double mutant, however, *Pprsl3 Pprsl4* plants developed a few normal caulonema filaments. The cells in these filaments were morphologically identical to the cells of WT caulonema (Fig. S3), confirming that the phenotype of *Pprsl3 Pprsl4* double mutants was caused by a reduction in the proportion of cells that undergo the chloronema-to-caulonema transition and not by a general protonema growth defect. *Pprsl3*, *Pprsl4*, *Pprsl5* and *Pprsl6* single mutants and *Pprsl5 Pprsl6* double mutants were phenotypically similar to WT protonemata (Fig. 3). We then constitutively expressed *PpRSL3* and *PpRSL4* under the control

of the CaMV 35S promoter. *35S:PpRSL3* plants were slightly smaller than WT and there was a strong reduction of caulonema development in *35S:PpRSL4* plants, which caused the development of very dense chloronema-rich protonemata (Fig. 3). These results indicate that ectopic expression or loss of function of *PpRSL3* and *PpRSL4* disrupts the chloronema-to-caulonema transition. Together, these data indicate that both *RSL* class I and class II genes control the chloronema-to-caulonema transition in the protonemata of *P. patens*, supporting the hypothesis that there is a functional RSL network in mosses.

The RSL networks have different topologies in *A. thaliana* and *P. patens*

In *A. thaliana*, a primary characteristic of the RSL network is the positive regulation of the *RSL* class II genes *AtRSL2* and *AtRSL4* by RSL class I proteins (26). To determine if the regulatory interactions between the two RSL classes are conserved between *P. patens* and *A. thaliana*, we measured the relative changes in the steady state levels of *RSL* mRNA in different *rsl* mutant backgrounds using quantitative RT-PCR. The levels of *A. thaliana RSL* class II mRNA were lower in *Atrhd6 Atrsl1* double mutants than in WT plants (Fig. 4A), indicating that RSL class I proteins positively regulate the transcription of all *RSL* class II genes. This conclusion is supported by the expression patterns of *RSL* genes in *A. thaliana* (Fig. 1C), which show that during differentiation of root hair cells *RSL* class I genes are expressed before *RSL* class II genes. Levels of *AtRSL3* and *AtRSL5* mRNA were also lower in the *Atrsl2 Atrsl4* double mutant (Fig. 4A), implying the existence of at least three levels of regulation in the *A. thaliana* RSL network: *RSL* class I genes positively regulate *AtRSL2* and *AtRSL4* which in turn positively regulate the transcription of *AtRSL3* and *AtRSL5*. In contrast, in the *P. patens Pprsl1 Pprsl2* double mutant only *PpRSL6* mRNA levels are altered compared with the WT (Fig. 4B). The levels of each *RSL* mRNA are also similar in WT, *Pprsl3 Pprsl4* and *Pprsl5 Pprsl6* double mutants (Fig. 4B). This indicates that the expression of the different *P. patens RSL* genes is largely independent of the activity of other

RSL proteins. Together, these results show that there are multi-level regulatory interactions between different *RSL* genes in *A. thaliana*, but fewer regulatory interactions between the *P. patens RSL* genes.

We also investigated the regulation of the RSL network by auxin, a key positive regulator of root hair development in angiosperms and the chloronema-to-caulonema transition in mosses (18, 32–34). Auxin positively regulates the expression of *RSL* class I genes in *P. patens* (24) (Fig. 4E) but had no effect on the expression of the *A. thaliana RSL* class I genes (Fig. 4D). However, the expression of *RSL* class II genes is highly responsive to auxin in both species. Exogenous auxin treatment increased the levels of *AtRSL4* and *AtRSL5* and reduced the levels of *AtRSL2* and *AtRSL3* mRNA in *A. thaliana* compared with untreated plants (Fig. 4D) (26). At low concentrations, exogenous auxin moderately increased the levels of all RSL class II mRNAs, whereas at higher concentrations it further increased the expression of *PpRSL6* and decreased the levels of *PpRSL3* (Fig. 4E). This means that auxin dynamically and strongly regulates the expression of *RSL* class II genes in both *A. thaliana* and *P. patens*. To confirm that endogenous auxin signalling modulates *RSL* gene expression, we compared *RSL* mRNA levels between wild type *P. patens* and *aux*/*iaa* mutants. Aux/IAA proteins are transcriptional repressors that mediate auxin signalling in land plants; a set of *aux/iaa P. patens* mutants are auxin-resistant and display a delay or arrest in the chloronema-to-caulonema transition similar to the phenotypes observed in *Pprsl* mutants (18, 35). In two of these *aux*/*iaa* mutants (*Ppiaa1A-113* and *Ppiaa2-183*), the mRNA levels of *RSL* class II genes was lower than in WT plants (Fig. 4C). Interestingly, a third *aux/iaa* mutant (*Ppiaa1B-112*) showed an inverse change in the levels of *PpRSL3* and *PpRSL4*, confirming our observation that auxin has a dynamic effect on the expression of *RSL* class II genes. Together these data demonstrate that auxin modulates the expression of *RSL* genes in *P. patens*.

We observed that the auxin-induced changes in the mRNA levels of *AtRSL4, AtRSL5*, *PpRSL3*, *PpRSL5* and *PpRSL6* were much larger in the *Atrhd6 Atrsl1* and *Pprsl1 Pprsl2* double mutants

than in wild type (Fig. 4D and F). This indicates that RSL class I proteins negatively regulate the transcriptional responses of these *RSL* class II genes to auxin. Conversely, while levels of *AtRSL2* and *AtRSL3* mRNAs decreased upon auxin treatment in WT plants, they increased upon auxin treatment in *Atrhd6 Atrsl1* double mutants (Fig. 4A and D). This indicates that RSL class I proteins are required for the auxin-induced repression of *AtRSL2* and *AtRSL3* in *A. thaliana*. The increase in steady state mRNA levels of *AtRSL2* and *AtRSL3* in auxin treated *Atrhd6 Atrsl1* double mutants is caused by the activity of *AtRSL4*; these genes were not induced when the *Atrhd6 Atrsl1 Atrsl4* triple mutant was treated with auxin (Fig. 4D). Together these data indicate that RSL class I proteins modulate the transcriptional responses of *RSL* class II genes to auxin.

These positive and negative interactions between *RSL* genes can be incorporated into a regulatory network with numerous feedforward loops, where auxin and *RSL* class I genes assume a central role in regulating the expression of *RSL* class II genes (Fig. 4G and H). The many differences in the *A. thaliana* and *P. patens* RSL network architecture demonstrate that extensive rearrangements of these regulatory networks have occurred since mosses and angiosperms last shared a common ancestor. A fundamental difference in the networks between the two species is that while the expression of individual *RSL* genes in *P. patens* is largely independent of the activity of other RSL proteins, the *A. thaliana* network is more hierarchical with both RSL class I and RSL class II proteins regulating the expression of *RSL* class II genes .

Discussion

Our functional and phylogenetic analyses suggest that RSL class I and class II transcription factors form a GRN kernel that evolved in the common ancestors of mosses and vascular plants (Fig. 5). The RSL kernel likely evolved as a mechanism controlling cell-type transitions in

aquatic charophyte algae or in the earliest land plants. After the divergence of mosses from other land plants, the RSL kernel controlled the differentiation of multicellular filamentous structures in mosses and was recruited to regulate the development of cellular projections from root epidermal cells (root hairs) in vascular plants. The architecture of the RSL kernel has changed since this divergence and became more hierarchical in the lineage that gave rise to *A. thaliana* than in the lineage that gave rise to *P. patens*. An alternative possibility is that the RSL kernel was already hierarchical in the first land plants and became reduced in the lineage that gave rise to modern mosses or, less likely, the RSL network evolved entirely independently in the two lineages.

Our expression analyses suggest that a close interaction of the RSL kernel with an auxin signalling 'plug-in' is a key structural feature of both the *P. patens* and *A. thaliana* GRNs (Fig. 4). Auxin is an important regulator of both protonema and root hair development and operates by regulating the expression of genes in the RSL network in both species. It is possible that other components of the higher-level GRN to which the RSL kernel belongs are partially conserved between mosses and angiosperms. We predict that there may be an overlap between the downstream differentiation gene batteries in both species. In *A. thaliana*, the RSL GRN regulates the transcription of a suite of genes that encode proteins involved in cell wall synthesis and modification, such as expansins, extensins and peroxidases (26, 36). Therefore we hypothesise that the RSL GRN controls the expansion of the root hair cell by regulating the expression of these growth effector proteins in root hairs. Given the conservation of elements of the RSL GRN among mosses and seed plants, it is conceivable that homologues of some of these effector proteins are required for the chloronema-to-caulonema transition in mosses. That is, the RSL GRN in *P. patens* may control the expression of growth effectors as in the root hair of *A. thaliana*. If this is the case, it raises interesting questions about how these different GRN components were recruited and re-assembled from the gametophyte to the sporophyte generation during the evolution of vascular plants.

The antiquity of the RSL kernel suggests that other plant regulatory networks may be derived from conserved GRN kernels that existed in the first land plants, almost 500 million years ago. Signalling pathways that can act as plugins of GRNs are conserved across land plants (18–20). Transcription factors such as KNOX/BEL and LEAFY proteins are conserved across plants (12–16, 37), but the networks in which these genes participate have not yet been described in early diverging groups of land plants. We predict that KNOX/BEL and LEAFY are also components of other plant kernels that control sporophyte development. If our hypotheses are correct, then the recruitment and modification of pre-Cambrian and/or Early Paleozoic gene regulatory networks were important evolutionary mechanisms in multicellular plants. This could have driven the generation of novel cell types and increased morphological diversity that occurred during the radiation of plants on land.

Materials and methods

The Columbia-0 (Col-0) wild type ecotype of *Arabidopsis thaliana* and the Gransden wild-type strain of *Physcomitrella patens* (Hedw.) Bruch and Schimp were used in this study. The lines *Atrhd6-3 Atrsl1-1*, *Atrsl2-1*, *Atrsl4-1*, *Atrsl2-1 Atrsl4-1*, *Atrhd6-3 Atrsl1-1 Atrsl4-1*, *Pprsl1-1 Pprsl2-1, Ppiaa1A-113, Ppiaa1B-112* and *Ppiaa2-183* were described previously (18, 23, 26). The T-DNA insertion *Atrsl3-1*: (GABI_422C06) was obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). *Atrsl5* was generated by introducing an artificial microRNA specific targeting to *AtRSL5* into the wild type. Two lines with significant repression for *RSL5* transcript were picked out for further phenotypic analysis. The generation of *Pprsl3*, *Pprsl4*, *Pprsl5*, *Pprsl6*, *Pprsl3 Pprsl4* and *Pprsl5 Pprsl6* mutants and of *Atrsl2 Atrsl4* plants transformed with *35S:AtRSL3, 35S:AtRSL5, 35S:PpRSL3, 35S:PpRSL4, 35S:PpRSL5* and 35S:*PpRSL6* constructs is described below.

See SI Methods for additional details.

Acknowledgments

We thank Jane Langdale, Monica Pernas and Sourav Datta for critical reading of the manuscript and Thomas Tam and Gloria Konstantoudaki for technical assistance.

Funding

NDP was supported by the Portuguese Fundação para a Ciência e a Tecnologia and the *EVO500* Advance Grant from the European Research Council to LD. KY was supported by a joint scholarship from the China Scholarship Council and the University of East Anglia with additional funding from The Human Frontiers in Science Program (RGP0012/2005-C). BM was funded by the EU-Marie Curie program (HPMF-CT-2002-01935) and a Natural Environmental Research Council (NERC) responsive mode grant NE/C510732/1 to LD. LD was also funded by the PLANTORIGINS Marie Curie Network. University of Oxford and John Innes Centre supported this research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contribution

Conceived and designed the experiments: NDP KY LD. Performed the experiments: NDP KY HB BC. Contributed reagents/materials/analysis tools: BM LD. Wrote the paper: NDP LD.

References

1. Carroll SB (2008) Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* 134:25–36.

2. Shubin N, Tabin C, Carroll S (2009) Deep homology and the origins of evolutionary novelty. *Nature* 457:818–823.

3. Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. *Science* 311:796–800.

4. De Robertis EM (2008) Evo-devo: variations on ancestral themes. *Cell* 132:185–95.

5. Erwin DH, Davidson EH (2009) The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* 10:141–8.

6. Gao F, Davidson EH (2008) Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. *Proc Natl Acad Sci USA* 105:6091–6.

7. Hinman VF, Nguyen AT, Cameron RA, Davidson EH (2003) Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. *Proc Natl Acad Sci USA* 100:13356–61.

8. Pimanda JE et al. (2007) Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc Natl Acad Sci USA* 104:17692.

9. Olson EN (2006) Gene Regulatory Networks in the Evolution and Development of the Heart. *Science* 313:1922–1927.

10. Mizutani CM, Bier E (2008) EvoD/Vo: the origins of BMP signalling in the neuroectoderm. *Nat Rev Genet* 9:663–677.

11. Melzer R, Wang Y-Q, Theißen G (2010) The naked and the dead: The ABCs of gymnosperm reproduction and the origin of the angiosperm flower. *Semin Cell Dev Biol* 21:118–128.

12. Harrison CJ et al. (2005) Independent recruitment of a conserved developmental mechanism during leaf evolution. *Nature* 434:509–514.

13. Tanahashi T, Sumikawa N, Kato M, Hasebe M (2005) Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens. *Development* 132:1727–1736.

14. Sakakibara K, Nishiyama T, Deguchi H, Hasebe M (2008) Class 1 KNOX genes are not involved in shoot development in the moss Physcomitrella patens but do function in sporophyte development. *Evol Dev* 10:555–566.

15. Sakakibara K et al. (2013) KNOX2 Genes Regulate the Haploid-to-Diploid Morphological Transition in Land Plants. *Science* 339:1067–1070.

16. Lee J-H, Lin H, Joo S, Goodenough U (2008) Early Sexual Origins of Homeoprotein Heterodimerization and Evolution of the Plant KNOX/BELL Family. *Cell* 133:829–840.

17. De Smet I et al. (2011) Unraveling the evolution of auxin signaling. *Plant Physiol* 155:209–21.

18. Prigge MJ, Lavy M, Ashton NW, Estelle M (2010) Physcomitrella patens Auxin-Resistant Mutants Affect Conserved Elements of an Auxin-Signaling Pathway. *Curr Biol* 20:1907–12.

19. Yasumura Y, Pierik R, Fricker MD, Voesenek L a CJ, Harberd NP (2012) Studies of Physcomitrella patens reveal that ethylene-mediated submergence responses arose relatively early in land-plant evolution. *Plant J*:947–959.

20. Khandelwal A et al. (2010) Role of ABA and ABI3 in desiccation tolerance. *Science* 327:546.

21. Cho SH, Coruh C, Axtell MJ (2012) miR156 and miR390 Regulate tasiRNA Accumulation and Developmental Timing in Physcomitrella patens. *Plant Cell* 24:4837–49.

22. Axtell MJ, Snyder JA, Bartel DP (2007) Common functions for diverse small RNAs of land plants. *Plant Cell* 19:1750–1769.

23. Menand B et al. (2007) An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316:1477–1480.

24. Jang G, Yi K, Pires ND, Menand B, Dolan L (2011) RSL genes are sufficient for rhizoid system development in early diverging land plants. *Development* 138:2273–81.

25. Jang G, Dolan L (2011) Auxin promotes the transition from chloronema to caulonema in moss protonema by positively regulating PpRSL1and PpRSL2 in Physcomitrella patens. *New Phytol* 192:319–327.

26. Yi K, Menand B, Bell E, Dolan L (2010) A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. *Nat Genet* 42:264–267.

27. Pires N, Dolan L (2010) Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. *Mol Biol Evol* 27:862–874.

28. Hedges S, Blair J, Venturi M, Shoe J (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol* 4:2.

29. Zimmer A et al. (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Mol Genet Genomics* 278:393–402.

30. Steemans P et al. (2009) Origin and Radiation of the Earliest Vascular Land Plants. *Science* 324:353.

31. Cove D (2005) The moss Physcomitrella patens. *Annu Rev Genet* 39:339–358.

32. Cormack RGH (1949) The development of root hairs in angiosperms. *Bot Rev* XV:583–612.

33. Johri MM, Desai S (1973) Auxin regulation of caulonema formation in moss protonema. *Nature New Biol* 245:223–224.

34. Jones AR et al. (2009) Auxin transport through non-hair cells sustains root-hair development. *Nat Cell Biol* 11:78–84.

35. Ashton NW, Grimsley NH, Cove DJ (1979) Analysis of gametophytic development in the moss, Physcomitrella patens, using auxin and cytokinin resistant mutants. *Planta* 144:427–435.

36. Velasquez SM et al. (2011) O-Glycosilated Cell Wall Proteins are Essential in Root Hair Growth. *Science* 332:1401–3.

37. Pires ND, Dolan L (2012) Morphological evolution in land plants: new designs with old genes. *Phil Trans R Soc B* 367:508–518.

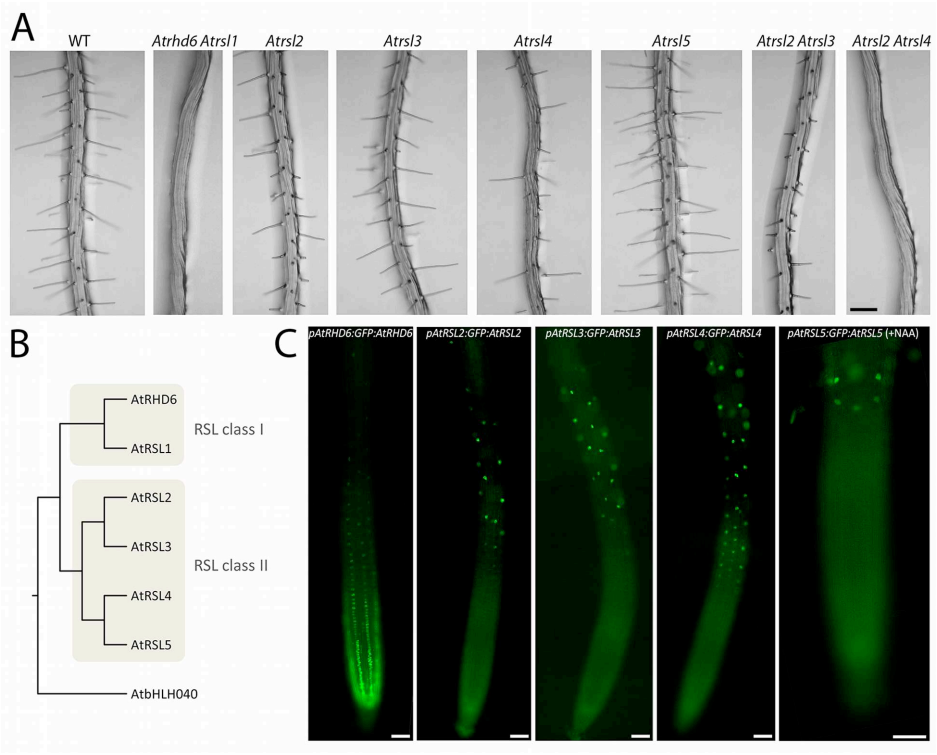

Fig. 1. RSL class I and RSL class II proteins control root hair development in *A. thaliana*. (**A**) Root hair phenotype of single and double mutants of *RSL* genes in *A. thaliana*. Scale bar: 200µm. (**B**) Maximum likelihood cladogram showing that the *A. thaliana RSL* genes fall into two classes. The tree was rooted with AtbHLH040 (27). (**C**) Promoter-GFP-protein constructs showing that RSL class I and class II proteins accumulate in the nuclei of root hair cells before and during root hair growth, respectively. GFP-AtRSL5 can only be detected after application of exogenous auxin. Scale bars: 50µm.

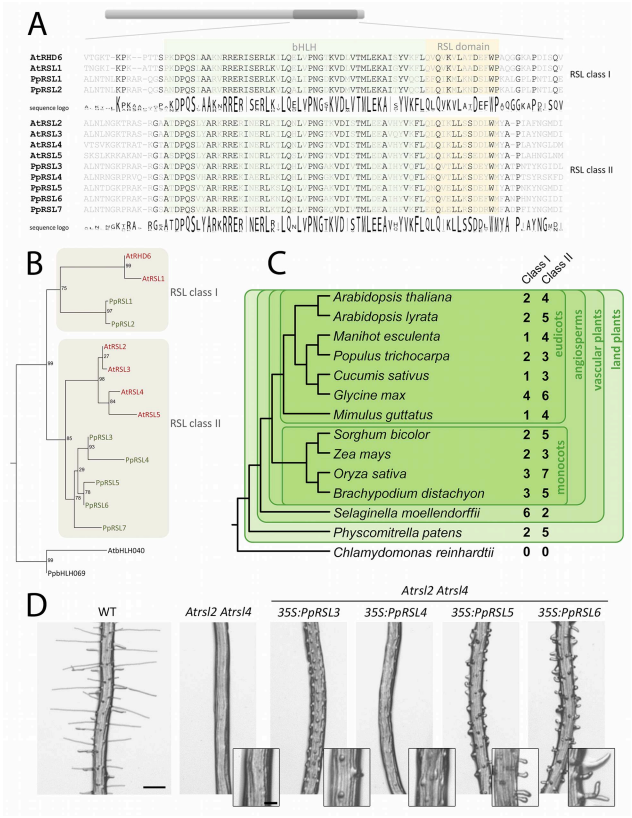

Fig. 2. RSL proteins are conserved across land plants. (**A**) Alignment of conserved regions of the *A. thaliana* and *P. patens* RSL proteins. The position of the bHLH RSL domains is indicated by coloured boxes; identical amino acids are represented in black. The sequence logos represent the multiple alignment of RSL class I and class II amino acid sequences from thirteen plant species (table S1); heights are proportional to sequence conservation in each position. (**B**) Maximum likelihood tree of *A. thaliana* (red) and *P. patens* (green) RSL proteins. The tree was based on the bHLH and RSL domains of the alignment shown in (**A**), together with the bHLH sequence of the outgroups AtbHLH040 and PpbHLH069 (27); aLRT support values are indicated in the nodes. See also Fig. S2. (**C**) Number of *RSL* class I and *RSL* class II genes in different plant species (see also table S1). (**D**) Six-day-old seedling roots of the *A. thaliana* wild type Col-0, *Atrsl2 Atrsl4* double mutant and *Atrsl2 Atrsl4* expressing the *P. patens* genes *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* under the control of the constitutive CaMV 35S promoter. Scale bar: 200µm in the main figure; 50µm in the close ups.

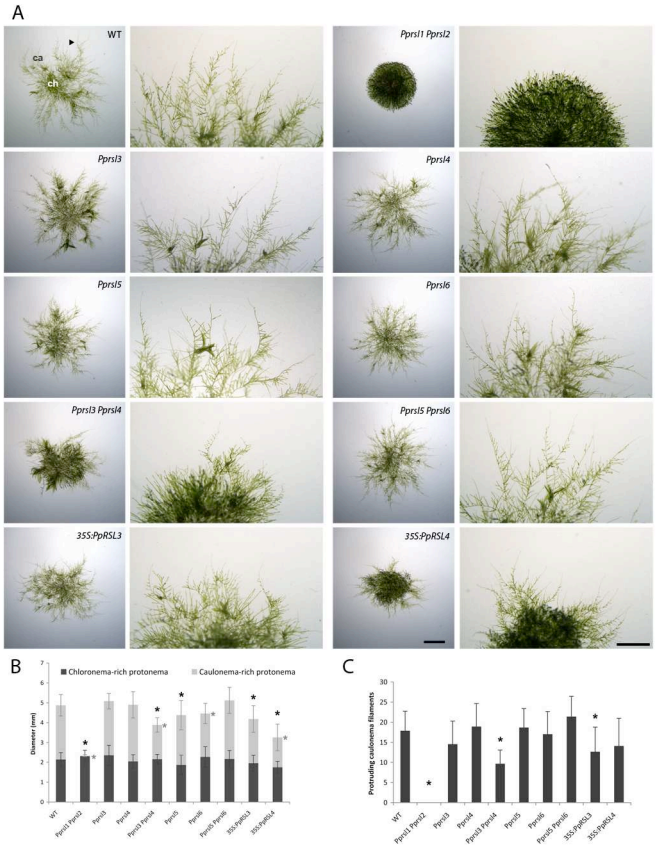

Fig. 3. RSL class II proteins control caulonema development in *P. patens*. (A) Protonemata of *Pprsl3*, *Pprsl4*, *Pprsl5*, *Pprsl6*, *Pprsl3 Pprsl4* and *Pprsl5 Pprsl6* loss of function mutants and constitutively expressed *PpRSL3* and *PpRSL4* genes in *P. patens*. Spores were germinated on minimal media for 3 weeks. Scale bars: 1mm. **(B)** Diameter of plants, shown as a stacked graph of the relative sizes of the inner chloronema-rich and the peripheral caulonema filament rich regions (mean ± SD, n=30) **(C)** Number of protruding caulonema filaments per plant (mean ± SD, n=30 plants). *Significantly different from WT (P < 0.01 with Bonferroni multiple comparison correction); grey asterisks refer to the caulonema-rich region alone.

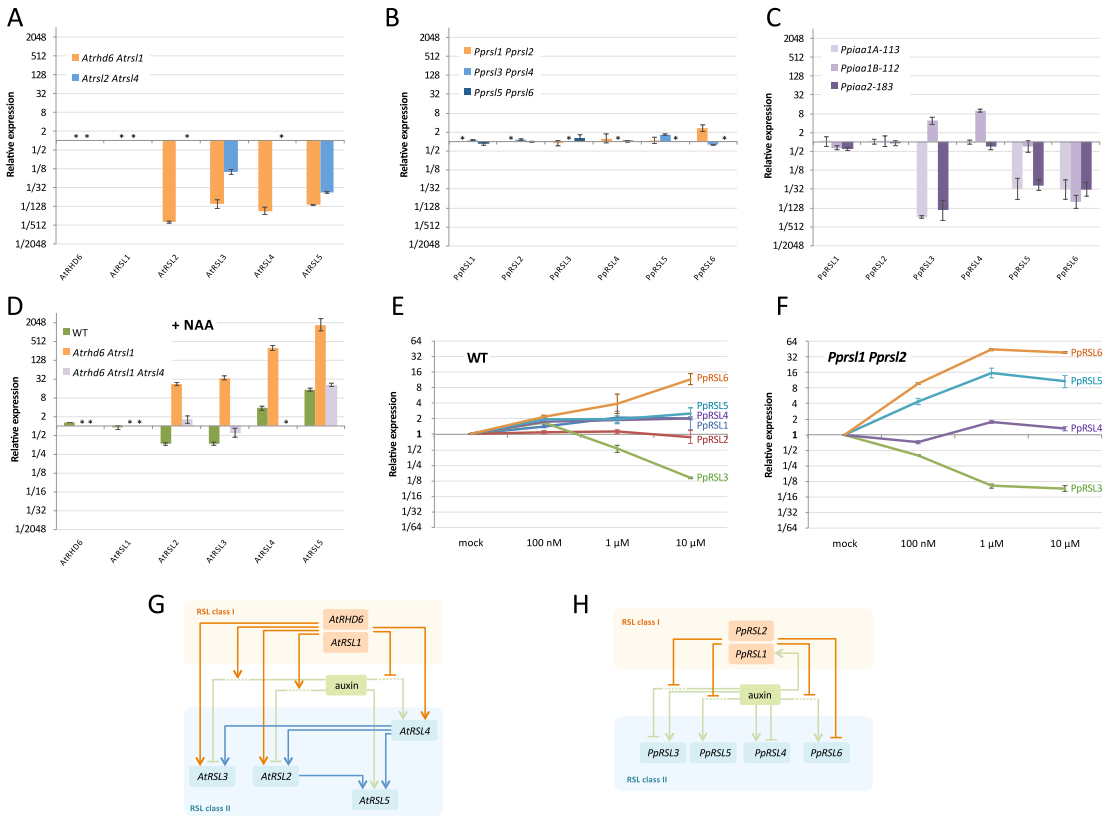

Fig. 4. RSL genes and auxin form regulatory networks in *P. patens* and *A. thaliana* (**A-F**) qRT-PCRs showing the relative expression level of *RSL* genes in different *A. thaliana* (**A, D**) and *P. patens* (**B, C, E, F**) mutant backgrounds and after NAA (1-naphthaleneacetic acid, a synthetic auxin) treatments. The expression levels are relative to WT (**A**-**C**) or untreated plants (**D-F**). The putative *PpRSL7* transcript was not detected. * absent/not determined. Bars represent the standard deviation of 3 independent replicates. (**G**, **H**) Schematic representation of the regulatory interactions between the different *RSL* class I genes (red), *RSL* class II genes (blue) and auxin (green) in *A. thaliana* (**G**) and *P. patens* (**H**).

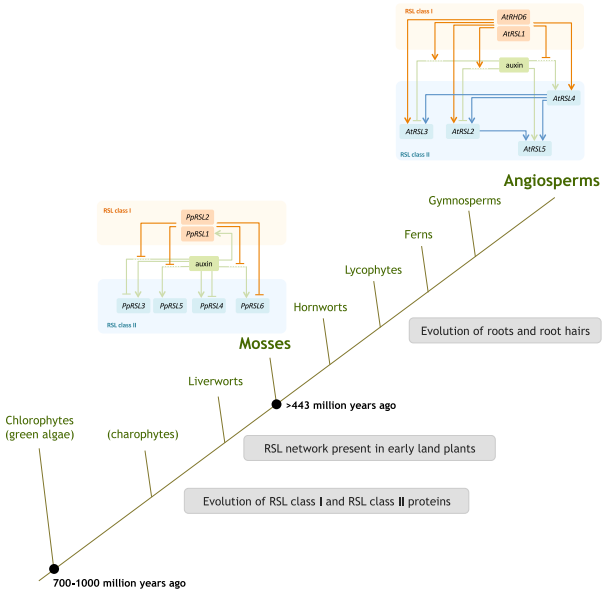

Fig. 5. Evolutionary history of the *RSL* network. *RSL* genes evolved in charophyte algae or in the first land plants, 500-1000 million years ago. An ancestral RSL kernel was present in early land plants and was conserved during land plant evolution. Later, during vascular plant evolution, the RSL network was recruited to control the development of cellular projections from root epidermal cells (root hairs).

Supporting Information

Pires et al. Recruitment and remodelling of an ancient gene regulatory network during land plant evolution

SI methods

Growth conditions

Arabidopsis thaliana seeds were sterilized in 5% bleach for 5 min and washed with sterilized water for 3 times. Seeds were sown in Petri-dishes containing sterilized media, Murashige and Skoog (MS) basal salt pH 5.7, 1% sucrose and 0.5% phytagel (Sigma, UK). Prior to germination, seeds were stratified at 4^{o}C for 4 days. The plants were grown vertically in the growth room with continual light at 25^{o}C. For auxin treatment, the seeds were geminated on the media described above but included 100 nM 1-naphthaleneacetic acid (NAA) and allowed to grow for 4 days. Moss spores were sterilized with 5% sodium hypochlorite for 10 min, washed five times with sterile water and stored at 4°C for several weeks. They were cultured under sterile conditions on minimal medium solidified with 0.8% agar [1], at 25°C with a 16 hours light/8 hours dark regime and a quantum irradiance of 40 µE $m^{-2}s^{-1}$. For the NAA treatments, spores were grown on minimal medium overlaid with cellophane disks for two weeks, after which the cellophane disks with protonema were transferred to media supplemented with 1µM NAA and incubated for a further week.

Constructs and plant transformation

The multisite gateway system was used to generate AtRSL3 and AtRSL5 GFP fusion vectors following the previously described procedure [2]. The primers used for the constructs were: RSL3PATTB4F: GGGGACAACTTTGTATAGAAAAGTTGAATAGTTAAACCTTACCCATCATCGG, RSL3PATTB1R: GGGGACTGCTTTTTTGTACAAACTTGTTTTGATCACTAAGCGACTTTAACAATAG, RSL3G3UtrATTB2F: GGGGACAGCTTTCTTGTACAAAGTGGAAATGGAAGCCATGGGAGAATG, and RSL3G3UtrATTB3R:

GGGGACAACTTTGTATAATAAAGTTGAGTCAGGCCTTTAGATAAGGTTTCTTAGT. Artificial microRNA precursors were conducted according to the protocol listed in: http://wmd2.weigelworld.org/themes/amiRNA/pics/Cloning_of_artificial_microRNAs.doc. PCR products of the amiRNA precursor fragment were digested with BamHI and KpnI and subcloned into the modified pCambia1300 with a CaMV 35S promoter and a terminator of Pea Rubisco small subunit. The primer pairs used for the constructs are RSL5A(I)miR-s: GATGACACTAAACCTAAGGCCCTTCTCTCTTTTGTATTCC, RSL5A(II)miR-a: GAAGGGCCTTAGGTTTAGTGTCATCAAAGAGAATCAATGA, RSL5A(III)miR*s: GAAGAGCCTTAGGTTAAGTGTCTTCACAGGTCGTGATATG, and RSL5A(IV)miR*a: GAAGACACTTAACCTAAGGCTCTTCTACATATATATTCCT. For the 35S:AtRSL3 and 35S:AtRSL5 constructs, the coding sequences of the genes were amplified from root cDNA with the primers: RSL3F(KpnI): CGGGGTACCATGGAAGCCATGGGAGAAT, RSL3R(BamHI): CGCGGATCCTCATCTGGTCAGTGCATTGAG for AtRSL3; RSL5F(KpnI): CGGGGTACCATGGAGAATGAAGCTTTTGTAGAT, RSL5R(BamHI): CGCGGATCCTTAAATAAGCCGAGACAAAAGATT for AtRSL5. The PCR fragments were digested with KpnI and BamHI, and cloned into the KpnI-BamHI sites of a modified pCambia1300 plasmid [2]. For the constitutive expression of moss genes in *Arabidopsis*, the coding sequences of *PpRSL3*, *PpRSL4*, *PpRSL5* and *PpRSL6* were amplified by RT-PCR and subcloned in pGEM-T Easy (Promega). A SmaI-PpRSL3-XhoI fragment was cloned into the EcoICRI-SalI sites of a modified pCambia1300 plasmid [2], generating the binary vector p35S:PpRSL3. A KpnI-PpRSL4-SalI fragment, a BamHI-PpRSL5-SalI fragment and a KpnI-PpRSL6-PstI fragment were cloned into the respective sites in the modified pCambia1300, generating the binary vectors p35S:PpRSL4, p35S:PpRSL5 and p35S:PpRSL6, respectively. All the binary vectors were transferred into the *Agrobacterium* strain GV3101. *Arabidopsis* plants were transformed using the floral dip method [3]. Transformants were selected on MS and agar plates containing 25 $\mu g.ml^{-1}$ hygromycin.

Physcomitrella knockout mutants were obtained by homologous recombination. The structure of each locus is shown in Fig. S4. Protonema phenotypes were confirmed in multiple independent lines. For the Pprsl3-ko construct, a 594bp fragment upstream and a 613bp fragment downstream of the *Pprsl3* bHLH coding sequence were amplified from genomic DNA by PCR (using the primers CCCCATGGCCATGGCCAACAGTCAGTAA / CCACTAGTCCACATTGCAGCTGCTTGAT and CCTCTAGACTTGGTACCAAATGGAGCTA / CCAAGCTTACATGCAGGATTTATCCTGG) and cloned into the NcoI-SpeI and XbaI-HindIII sites of pBHSNR [4], respectively. For the PpRSL4-ko construct, a 734bp fragment upstream and a 987bp fragment downstream of the *PpRSL4* coding sequence were amplified from genomic DNA by PCR (using the primers GCCTCTAGATCGTGGCTTTCTTTCAGGTG / GGCTCGAGTCTTCTCAACGGGTGCTTCA and CGGACTAGTGGTACCGCCGAAATCTACCA / CCGACGCGTTCAGCAGTGCAAACTTTGGTT) and cloned into the XbaI-XhoI and SpeI-MluI sites of pBNRF [4], respectively. For the PpRSL5-ko construct, a 698bp fragment upstream and a 710bp fragment downstream of the *PpRSL5* coding sequence were amplified from genomic DNA by PCR (using the primers CGCCGTCGACATTCCATTTTCTGGCGCTTG / CCGGATCCTGCTGTCACATGAAATGAGTTG and GCCTCTAGAAACCACGCCGCTCATAATTT / CCGACGCGTTGTTGCCTCTTGTCGTGTGT) and cloned into the SalI-BamHI and SpeI-MluI sites of pBHSNR, respectively. For the PpRSL6-ko construct, a 712bp fragment upstream and a 602bp fragment downstream of the *PpRSL6* coding sequence were amplified from genomic DNA by PCR (using the primers CTTGGTAACGTGGACAGCTCGAT / GACGCTACTCTGCGGTTAGTCAGG and CGTGTGTAACAGCCCCACCAG / CATCAACCAAAATGTATTTCAATGG) and cloned into the SalI-HindIII and MluI-NsiI sites of pBZRF (gift from Fabien Nogue, Versailles). For the constitutive expression of genes in *Physcomitrella*, a 2973bp fragment from pGWB2 (Tsuyoshi Nakagawa, Japan), carrying a 35S promoter-attR1-CmR-ccdB-attR2 cassette, was amplified by PCR (using the primers CCGCGGCCGCGTTGAATGTCGCCCTTTTGT and GCACTAGTCGGAAATTCCTCTCCTGTCA) and cloned into the NotI-SpeI sites of p108-lox-35Snpt-

lox, carrying the 108 locus of *Physcomitrella* (gift from Prof. Pierre Goloubinoff, Lausanne, Switzerland), generating the moss transformation vector p108GW35S. The coding sequences of *PpRSL3* and *PpRSL4* were amplified from protonema cDNA by RT-PCR, subcloned into pCR8/GW/TOPO-TA (Invitrogen) and then cloned into p108GW35S by LR reaction using the Gateway LR Clonase II enzyme mix (Invitrogen), generating the moss transformation constructs p108oxRSL3 and p108oxRSL4. The overexpression of PpRSL3 and PpRSL4 was confirmed by qRT-PCR.

Moss protoplast isolation and polyethylene glycol-mediated transformation were performed as described previously [5]. Selection for antibiotic-resistant transformants was done using 50 $\mu g.ml^{-1}$ G418 disulfate, 25 $\mu g.ml^{-1}$ hygromycin B or 50 $\mu g.ml^{-1}$ zeocin (Invitrogen). Stable transformants were confirmed by PCR and Southern blot. Primers used for these PCRs were: (Pprsl3) p1-AGTCGCCTTCCTCTCCTCTC, p2-CGGTGAGTTCAGGCTTTTTC, p3-TCCGAGGGCAAAGAAATAGA, p4-TCAGTTGCCTTCTTGTGTGC, p5-CGACTGATCCGCAGAGTGTA, p6-AATGTCTTCAAGCGCTCGTT; (Pprsl4) p1-ATGGATGGCTGAGGTGTTGT, p2-TTGCTTTGAAGACGTGGTTG, p3-CTTCGACGGATCTCGACCT, p4-GGCATGAGCTACCAAAAGGTG, p5-TCTTCGGGGATCTAGCTGTC, p6-TGGCGTACATCCAATACTCG; (Pprsl5) p1-CAAAGAATCTGAACGGCTCAA, p2-GGTGGAGCTCGGTACCATAA, p3-CCGGCCAGATCTATAACTTCG, p4-GTGGAGCTAGCCGCAGATG, p5-TCATGCATCGAAACCTCGTC, p6-TCTCCTCAAGTTCAAGAGGGTGT; (Pprsl6) p1-ACAGCTTCGGCCTTTCACTA p2-CGTGGGATCCTCTAGAGTCG, p3-GCCGGCCAGATCTATAACTTC, p4-TTTTATCTCCGATTCTTATGTCTAAGT, p5-GCGGTCCTACTTCCATTCTG, p6-ATGCCGTTGTAGTTGTGTGG. For Southern blots analysis, genomic DNA was extracted using the Nucleon Phytopure Genomic DNA Extraction Kit (GE Healthcare) and 1 µg DNA was digested overnight with 100U of the appropriate restriction enzyme. After electrophoresis, the DNA was transferred to a positively charged nylon membrane. Hybridization and detection were performed as described in the Roche DIG Application Manual: using PCR DIG Probe Synthesis

Kit, DIG Easy Hyb Granules, DIG Luminescent Detection Kit and the Lumi-Film Chemiluminescent Detection Film (Roche). The primers used for the PCR labelling of the probes were ATCCGGTCGGCATCTACTCT and TGTAGGAGGGCGTGGATATG (Hyg probe), TGAATGAACTGCAGGACGAG and AATATCACGGGTAGCCAACG (nptII probe), GACTAAACCTGGAGCCCAGAC and GAACTAGTGGATCCCCGTCA (Zeo probe)

Sequence retrieval and phylogenetic analyses

An initial alignment of conserved amino acid sequences from RSL class I and RSL class II proteins [6] was used to build an RSL specific pHMM with the program hmmbuild [7]. This pHMM was used to identify *RSL* coding sequences with the program hmmsearch [7] in the gene model databases of *Brachypodium distachyon* [8], *Chlamydomonas reinhardtii* v4.0 [9], *Mimulus guttatus* v1.0, *Physcomitrella patens* [10], *Populus trichocarpa* v1.1 [11], *Sorghum bicolor* [12] and *Selaginella moellendorffii* v1.0 [13] from the DOE Joint Genome Institute (http://www.jgi.doe.gov/); *Arabidopsis lyrata*, *Cucumis sativus*, *Glycine max* [14] and *Manihot esculenta* v1.1 from Phytozome 5.0 (http://www.phytozome.net/); *Arabidopsis thaliana* from The Arabidopsis Information Resource (http://www.arabidopsis.org/); *Oryza sativa* 6.1 from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/) and *Zea mays* from the Maize Genome Sequencing Project (http://www.maizesequence.org/). Amino acid sequences of the bHLH and RSL domains were manually aligned and ML analyses were performed with PhyML version 3.0.1 [15] using a JTT model of amino acid substitution, an estimated gamma distribution parameter and a SH-like aLRT. Trees were visualised using Figtree (http://tree.bio.ed.ac.uk/software/figtree/). Logos of amino acid sequence alignments were created with WebLogo (http://weblogo.berkeley.edu/).

Microscopy

Roots and protonema were imaged with a Leica DFC310 FX camera mounted on a Leica M165 FC stereo microscope. GFP was imaged with a Hamamatsu Orca HQ cooled CCD digital camera (Hamamatsu Photonics) mounted on an epifluorescence light microscope Nikon Eclipse E600 with ×20/0.7 Nikon Plan Apo objectives using a normal fluorescein isothiocyanate filter. For the visualization of individual protonema cells, protonema filaments were dissected from 21 days old protonema (growing on minimal media overlaid with cellophane disks), mounted on a 100 $\mu g.ml^{-1}$ Calcofluor White solution (Sigma) and imaged with a Retiga EXi CCD camera (Qimaging) mounted on a Olympus BX50 microscope with an excitation filter of 330-385nm and a barrier filter of 420-385nm. Images were post-processed with ImageJ (http://rsb.info.nih.gov/ij/) and Adobe Photoshop.

qRT-PCR

Total RNA was isolated from frozen plant tissue with the RNeasy Plant Mini Kit (Qiagen) and reverse transcribed using the SuperScript III First-Strand Synthesis System for RT-PCR using oligo(dT) (Invitrogen). qRT-PCR analyses were performed with the DyNAmo SYBR Green qPCR kit (Finnzymes, Espoo, Finland) in a MJ Opticon (MJ Research, MA, USA) thermal cycler and with the SYBR Green PCR Master Mix (Applied Biosystems) in an Applied Biosystems 7300 Real-Time PCR System. Specific primers were designed to generate PCR products between 100 and 200 bp: AtRHD6: CCTAAATCCGCTGGAAACAA and CTCTTCGATTCTTGGCTGCT; AtRSL1: CCCTAAACTGGCTGGCAATA and TCTTGGCTGCTAGGCTTTGT; AtRSL2: CCCCAATGGAACAAAGGTC and TCTCGGTGAGCTGAGACCAA; AtRSL3: GGAGCCAGAAATGCGTAGAG and GTCTCCACCGTTTGATTCGT; AtRSL4: GTGCCAAACGGGACAAAAGT and TTGTGATGGAACCCCATGTC; AtRSL5: GCAGGAACTTCACGTAATGGA and TATACGCTAGGAAACGAAGAGAAA; AtEF1α: GGTGGTGGCATCCATCTTGTTACA and TGAGCACGCTCTTCTTGCTTTCA; PpRSL1: GTGTCCTATTCCGAGGACCA and

CAACCTTGAGGTCCCAAAAA; PpRSL2: AGGATAAGTGAGCGGCTGAA and CCTTGGGCCAGATAGAATCA; PpRSL3: AACGAGCGCTTGAAGACATT and TTCAACAGCGTCACTTGGAG; PpRSL4: CGACTGATCCGCAGAGTGTA and GGTTACGATGTCCACCTTCC; PpRSL5: GCAACCGATCCTCAGAGTGT and TCAACCTTGGCTCCATTAGG; PpRSL6: AAATCTCGTGCCAAATGGAG and CATCCAGAACTCGTCGGATT; PpGAPDH: CTTGAGAAGCCTGCCTCCTA and TGCTGTCGGTAATGAAGTCG; PpEF1a: GGATCTTGTCGGGGTTGTAA and TTTCACCTTGGGAGTGAAGC. Relative expression levels were calculated by the ΔΔCt method, using the Elongation Factor 1α (At5g60390) transcript for normalisation in *Arabidopsis thaliana* and GAPDH (PHYPADRAFT_226280) and Elongation Factor 1α (PHYPADRAFT_158916) in *Physcomitrella patens*.

SI references

1. Ashton NW, Cove DJ, Featherstone DR (1979) The isolation and physiological analysis of mutants of the moss, Physcomitrella patens, which over-produce gametophores. Planta 144: 437-442.

2. Yi K, Menand B, Bell E, Dolan L (2010) A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. Nat Genet 42: 264-267.

3. Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16: 735-743. doi:10.1046/j.1365-313x.1998.00343.x

4. Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E, et al. (2007) An ancient mechanism controls the development of cells with a rooting function in land plants. Science 316: 1477-1480.

5. Schaefer DG, Zrÿd J-P (1997) Efficient gene targeting in the moss Physcomitrella patens. Plant J 11: 1195–1206.

6. Pires N, Dolan L (2010) Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants. Mol Biol Evol 27: 862-874. doi:10.1093/molbev/msp288

7. Eddy SR (1998) Profile Hidden Markov Models. Bioinformatics 14: 755 - 63.

8. The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463: 763-768.

9. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. Science 318: 245-250. doi:10.1126/science.1143609

10. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319: 64-69.

11. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313: 1596-604. doi:10.1126/science.1128691

12. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, et al. (2009) The Sorghum bicolor genome and the diversification of grasses. Nature 457: 551-556.

13. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, et al. (2011) The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. Science 332: 960-3. doi:10.1126/science.1203810

14. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178-183.

15. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704.

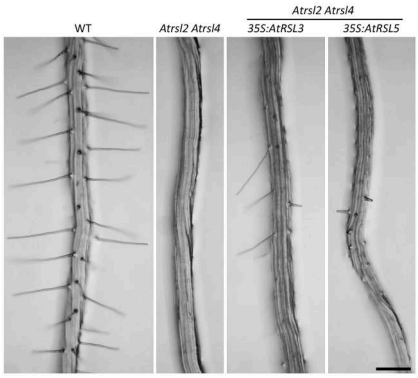

Fig. S1. Roots of the *Arabidopsis thaliana* wild type Col-0, *Atrsl2 Atrsl4* double mutant and *Atrsl2 Atrsl4* plants expressing *AtRSL3* and *AtRSL5* under the control of the constitutive CaMV 35S promoter. The scale bar indicates 200µm.

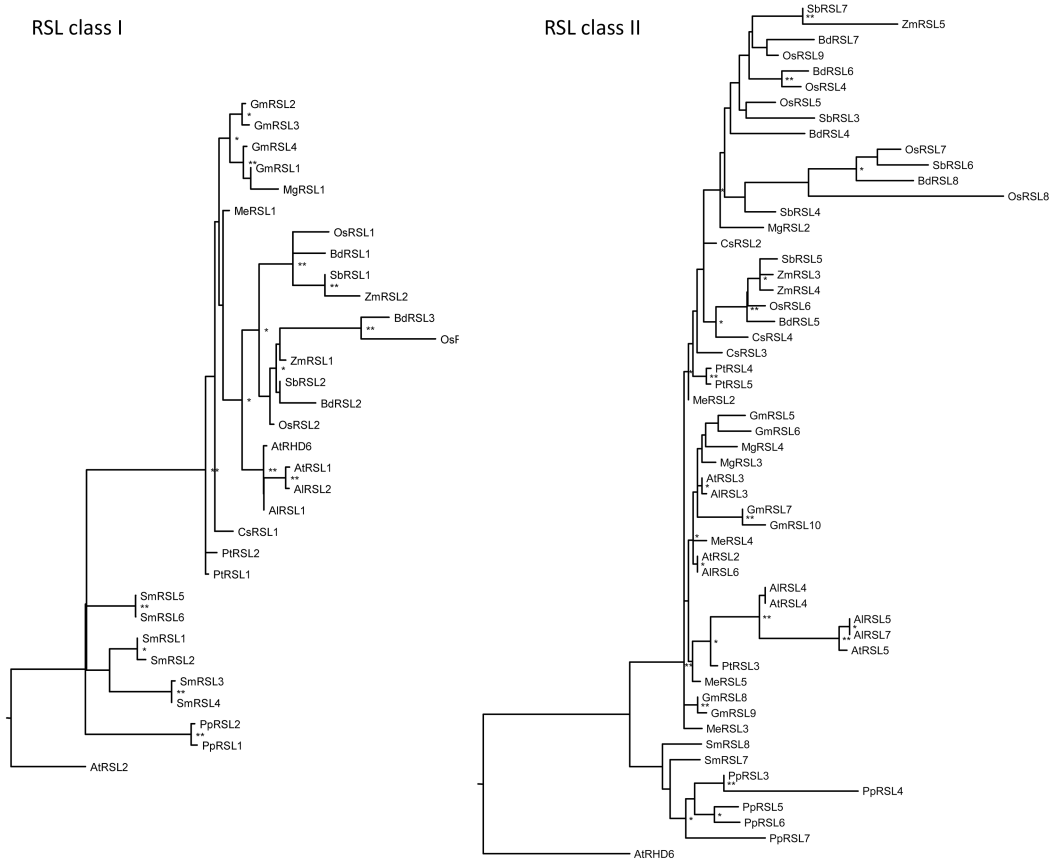

Fig S2. ML phylogenetic trees of RSL class I and RSL class II genes from *Arabidopsis lyrata* (Al), *Arabidopsis thaliana* (At), *Brachypodium distachyon* (Bd), *Cucumis sativus* (Cs), *Glycine max* (Gm), *Manihot esculenta* (Me), *Mimulus guttatus* (Mg), *Oryza sativa* (Os), *Physcomitrella patens* (Pp), *Populus trichocarpa* (Pt), *Sorghum bicolor* (Sb), *Selaginella moellendorffii* (Sm) and

Zea mays (Zm). The RSL class I tree was rooted with AtRSL2 and the RSL class II tree was rooted with AtRHD6. aLRT support values are indicated in the nodes (* > 0.8; ** > 0.9).

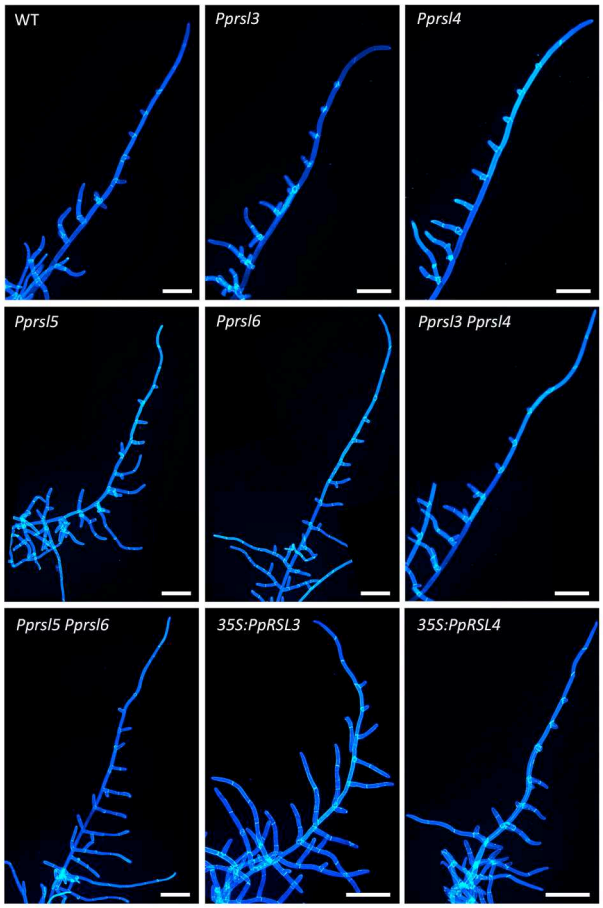

Fig. S3. Filaments protruding from 21 days old *Physcomitrella patens* protonema grown on minimal media overlaid with cellophane disks were isolated, stained with Calcofluor White and observed in an epifluorescence microscope. Caulonema cells can be distinguished from chloronema cells by their larger length and oblique cross cells walls. Scale bars indicate 200µm.

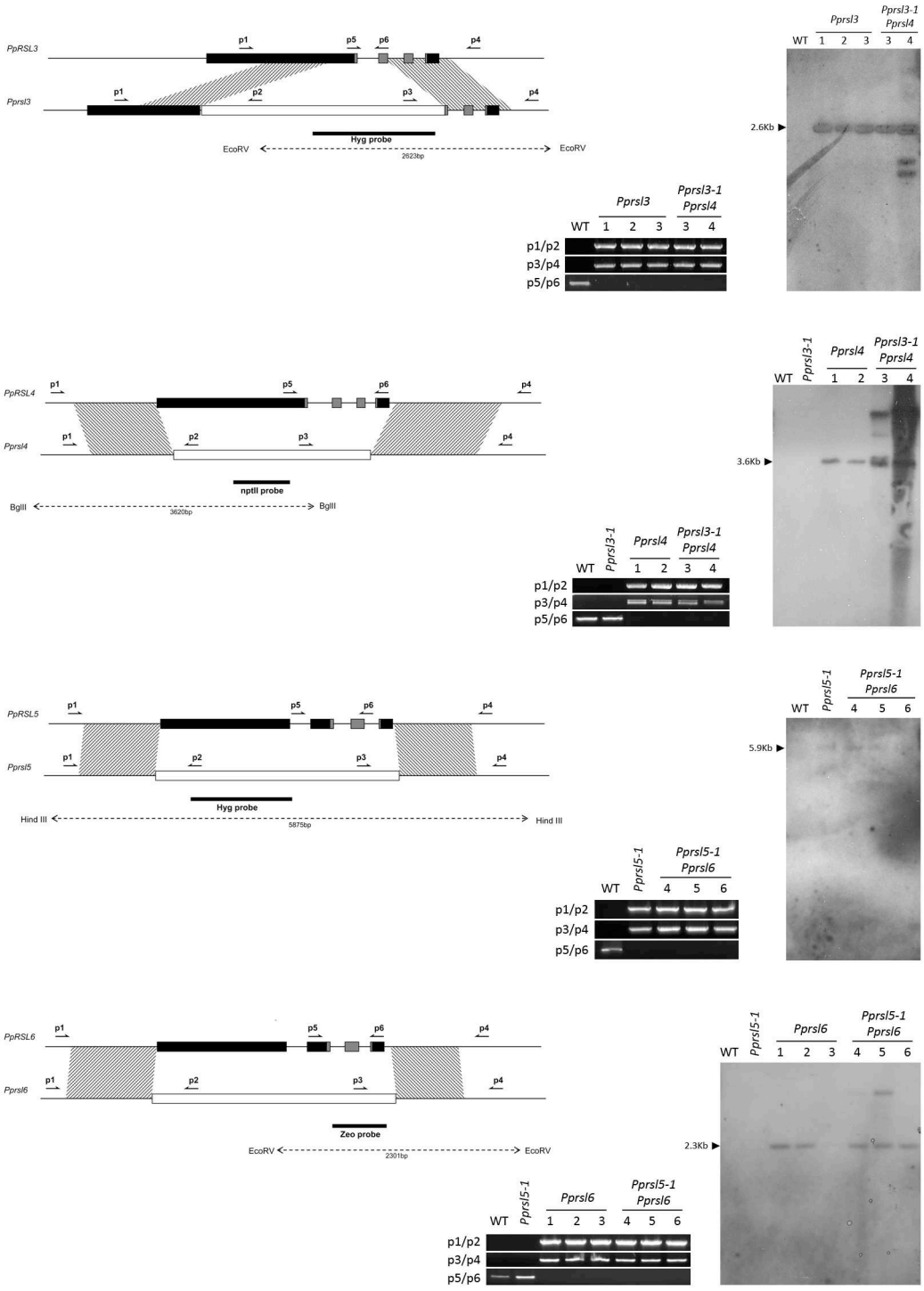

Fig. S4. Construction of *P. patens* loss of function mutants by homologous recombination. The structure of each locus is indicated. The dark boxes correspond to exons, with the grey boxes indicating the position of the bHLH domain; the white box indicates the cassette that confers resistance to hygromycin, G418 or zeocin. The dashed region indicates the position of the genomic region used for the homologous recombination. The harpoons indicate the location of

the primers (p1-p6) used in the PCRs to confirm the recombination events. The black bar indicates the location of the probe used in the Southern blots to confirm the stable integration of the constructs in correct position. The location of the corresponding restriction sites is indicated by a dashed arrow.

Table S1. RSL sequences used for the phylogenetic analyses.